54 p.

MEMORANDUM
RM-4019-NASA OTS:
FEBRUARY 1964

UNPUBLISHED PRELIMINARY DATA

ON GENERAL LEAST SQUARES ESTIMATION OF SOLAR SYSTEM CONSTANTS

R. T. Gabler and Albert Madansky

PREPARED FOR:

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

The RAND Corporation
SANTA MONICA • CALIFORNIA

MEMORANDUM
RM-4019-NASA
FEBRUARY 1964

ON GENERAL LEAST SQUARES ESTIMATION OF SOLAR SYSTEM CONSTANTS

R. T. Gabler and Albert Madansky

This research is sponsored by the National Aeronautics and Space Administration under Contract No. NASr-21. This report does not necessarily represent the views of the National Aeronautics and Space Administration.

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA

PREFACE

Various NASA groups and contractors are faced with the problem of adopting values for such solar system constants as the astronomical unit (in effect, the mean earth-sun distance in kilometers).

This quantity, together with other related constants such as planetary masses, is needed for accurate and efficient guidance of space craft. The need therefore exists for a systematic method of standardization for choosing values for such constants and for improving estimates of them as new data become available. This Memorandum is in essence an exposition of the application of the theory of measurement errors and statistical estimation precedures to such problems. It deals specifically with the general least squares estimation method and illustrates how this method can be effectively applied to determine the "best values" for solar system constants, to estimate their variance, and combine new measurements with what is already known.

The Memorandum should be of value to scientists working in the field of guidance and navigation of space vehicles, as well as those interested in celestial mechanics. Since the methods outlined are quite general, and not limited to the specific field of astrodynamics, the paper will also be of interest to scientists concerned with determining physical constants from theoretically related experiments.

SUMMARY

17225

This Memorandum reviews the statistical theory of parameter and variance estimation and its application to the problem of estimating solar system constants. The usual statistical methods for treating experimental data are discussed, with particular attention to the concepts of random error and systematic error. The maximum likelihood method for parameter estimation, which is a preferred statistical estimation method, is shown to coincide with the general least squares method when errors are distributed as the multivariate normal distribution, which is usually true to a satisfactory degree of approximation. The general least squares method is therefore of great potential utility and it is discussed in detail, especially with regard to methods for combining results from several related experiments. In appendices the theory is applied to several illustrative examples and to a specific experiment that has been suggested for determining the astronomical unit. Author

ACKNOWLEDGMENTS

During the course of writing this Memorandum, encouragement and helpful discussion has come from G.F. Schilling, A.G. Wilson, M.H. Davis, D.W. Stebbins and D.S. Wilson, of RAND. In the same category are discussions with Samuel Herrick and Gerald de Vaucouleurs and Nelson Wax, RAND consultants, as well as E.M. Boughton and L. Wong, of Space Technology Laboratories. Helpful discussion on certain aspects of mathematical statistics has also come from Allan Marcus, RAND consultant and graduate student at University of California, Berkeley.

The work of Louise Kern in the computation of and checking of formulae in the examples is gratefully acknowledged. The help of W.L. Sibley in locating and interpreting the appropriate numerical computing routines for use in the examples is very much appreciated. Input data for example 3 in Appendix I was furnished by Dean Jamison, of RAND.

CONTENTS

I. INTRODUCTION

In the analyses, design, and implementation of space experiments various groups are faced with the problem of obtaining or establishing values for repeatedly used constants and parameters. There is need not only for precise numerical values, but also a need for a measure of probable accuracies for these values.

Through observations over centuries, astronomers have established a mathematical model in which most of the solar system dynamical constants and parameters are related. They have adopted preferred values when it seemed appropriate and desirable.[1-5] The present capability of sending vehicles into space increases both the opportunity for more observations and more types of observation and also the need for better estimates of parameters. As more and better observations are made, it is possible in principle to make new and more accurate estimates for these solar system constants. Occasionally one may need to revise a mathematical model and introduce and estimate new parameters.

Often details of the methods by which new values for constants are estimated are not fully disclosed by researchers. This makes it difficult to assess the validity of the estimated value of the constant, and makes it impossible to combine new data with this estimate in a systematic way to arrive at a better estimate of the constant.
In many cases, even when a well-known and precisely defined

estimation method is used, the method for determining the probable error or standard deviation of the estimate is obscure, and statements about its accuracy may be ambiguous. The purpose of this Memorandum is to describe the general least squares estimation method, in the hope of promoting standardization in the method of solar system constant estimation, and uniformity in the statement of results of the estimation. By this method, as new data are obtained, estimates of constants may be updated in a systematic manner, and standard deviations of the updated estimates are obtained in the process. The method described here is by no means original, but some aspects of the method may appear novel even to least-squares practitioners.

The applicable statistical theory is reviewed and examples of the applications are worked out. No apology is made for the abbreviated presentation of certain ideas and methods from statistics, since the objective here is to focus on least squares theory rather than on the more general field of multivariate statistical analysis.

A formulation of the general least squares problem using matrix methods is presented in Arley and Buch[6] and this has been extended by D.C. Brown.[7] In a recent book, Linnik[8] develops the method using matrices. Geodesists have made rather extensive use of general least squares methods, and a pertinent example of the use of procedures delineated by Arley and Buch[6] and D.C. Brown[7] is available in a report by W.M. Kaula and Irene Fischer,[9] concerned with the use of geodetic and astronomic measurements to obtain a World Geodetic System.

II. ERROR THEORY

No matter how precisely one attempts to measure physical quantities, the observed measurement differs from the true magnitude of the quantity measured by what is called a "measurement error." The classical theory of measurement errors is concerned with only one kind, random errors. Two other groups, sometimes labeled coarse errors and systematic errors, must be removed or their effects must be small if a statistical analysis of the measurements is to yield a good estimate of the quantities measured.

Coarse errors usually result from lack of care on the part of an observer, and because they become conspicuous when data are displayed in one way or another they may be eliminated by discarding obviously erroneous observations.

The following quotation from Arley and Buch[6] is an apt description of systematic error:

"Systematic errors are errors due to one or a few definite causes acting according to a definite law and, as a rule, in one definite direction. If a measurement is repeated under constant conditions, the same systematic errors will occur. Consequently, in contrast to the coarse errors, systematic errors will not show up in any disagreement among different results, but only displace them by a constant amount. However, if the laws governing systematic errors are known, these errors can be calculated and treated as corrections to the values measured. Most systematic errors are caused by the instruments." Unfortunately, systematic errors are not as easily discovered or

eliminated as are coarse errors.

To quote again from Ref. 6, "Random errors are all the other errors which do not show any regularities or the regularities of which we do not know. Sometimes the word error is applied only to systematic errors and the word uncertainties to the random errors. In general, it is a characteristic feature of random errors, in contrast to systematic errors, that positive and negative values are equally probable. However, errors having skew distributions may be found. These are the so-called one-sided errors. By closer investigations, one-sided errors often turn out to be systematic errors. As an example of a one-sided error, we may mention the curvature in the axes of optical instruments.

"The distinction between the various groups of errors is, however, not sharp. By closer investigation, some of the random errors may show regularities. Thus an error we have previously classified as random may later turn out to be systematic. Casually there may also appear a particularly large error which may be mistaken for a coarse error. Thus in practice, it is not always easy to judge whether or not a measurement which deviates conspicuously from the other measurements should be rejected."

In the theory of errors, the probability distribution of the measurement errors is nearly always assumed to be the normal or Gaussian distribution function. If x is an observation on the parameter μ, then its probability density function is given by the normal density function

$$\phi(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\tfrac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

where σ is the standard deviation of x.

In experimentation one usually deals not with just a single observation on a single parameter, but with replicated observations on many parameters. To deal with this more general case, the probability distribution of the measurements is usually assumed to be the multivariate normal distribution. If $(x_1, \ldots, x_n)$ are observations on $(\mu_1, \ldots, \mu_n)$ where the μ_i's need not be distinct, then the probability density function of the vector* $x = (x_1, \ldots, x_n)$ is given by

$$\phi(x) = \frac{|A|^{-\frac{1}{2}}}{(2\pi)^{n/2}} \exp\left[-\tfrac{1}{2}(x-\mu)\, A^{-1}(x-\mu)\right]$$

where $\mu = (\mu_1, \ldots, \mu_n)$, and the matrix **A**, called the covariance matrix of x, has as its i-th diagonal element the variance of x_i and as (i,j)-th element, $i \neq j$, the covariance of x_i and x_j.

Some particular special cases are worth noting explicitly. When $\mu_1 = \mu_2 = \ldots = \mu_n = \mu$, say, then $x_1, \ldots, x_n$ are replicated observations on μ, though possibly with different variances and possibly correlated. When **A** is a diagonal matrix, then $x_1, \ldots, x_n$ are uncorrelated and so are independent univariate normal variables, but they may have different variances. When $A = \sigma^2 I$, where I is the identity matrix and σ^2 is a scalar, $x_1, \ldots, x_n$ are independent and identically distributed univariate normal variables.

---

*Throughout this paper we shall write all our vectors as row vectors, so that the transpose (denoted by a prime) of a vector will be a column vector.

III. MAXIMUM LIKELIHOOD ESTIMATION

The purpose of most scientific experiments is the estimation of parameters. The major role of the statistician in these experiments is to provide satisfactory estimation procedures. The point of view of the statistician, and the problems that beset him, are well stated in the following quotation from Cramer.[10]

"Suppose that we are given a sample from a population,* the distribution of which has a known mathematical form, but involves a certain number of unknown parameters. There will then always be an infinite number of functions of the sample values that might be proposed as estimates of the parameters. The following question then arises: How should we best use the data to form estimates? This question immediately raises another: What do we mean by the 'best' estimates?

"We might be tempted to answer that, evidently, the best estimate is the estimate falling nearest to the true value of the parameter to be estimated. However, it must be borne in mind that every estimate is a function of the sample values and is thus to be regarded as an observed value of a certain random variable. Consequently, we have no means of predicting the individual value assumed by the estimate in a given particular case, so that the goodness of an estimate cannot be judged from individual values, but only from the distribution of the

*Sample is a set of observations, and population is a very large set which has the assumed characteristic distribution.

values which it will assume in the long run; i.e., from its sampling distribution. When the great bulk of the mass in this distribution is concentrated in some small neighborhood of the true value there is a great probability that the estimate will differ from the true value by a small quantity. From this point of view, an estimate will be 'better' in the same measure as its sampling distribution shows a greater concentration about the true value and the above question may be expressed in the following more precise form: How should we use our data in order to obtain estimates of maximum concentration?

"The concentration (or the complementary property: the dispersion) of a distribution may be measured in various ways, and the choice between various measures is to a great extent arbitrary. The same arbitrariness will, of course, appear in the choice between various estimates. Any measure of dispersion corresponds to a definition of the 'best' estimate, viz. the estimate that renders the dispersion as expressed by this particular measure as small as possible."

Many criteria may be used in evaluating a method of estimation, but it is clear from the above that an important property of an estimator is that its sampling distribution have maximum concentration or minimum variance. It can be shown that for large sample sizes estimators based on the method of maximum likelihood have this desired property. This property, plus the fact that these estimators are fairly easy to construct, is mainly responsible for the attention and use which the maximum likelihood method has received. To use this estimation method, the probability density function of the sample (involving the parameters to be estimated), the likelihood function, must be specified.

In the theory of errors, the general form of the likelihood function is specified, as indicated above, as the multivariate normal density function. The exact form of the multivariate normal density function to be used depends on particular detailed aspects of the experiment. We distinguish two main cases--that of direct measurement and that of indirect measurement on a set of m parameters to be estimated.

The simplest case of direct measurement is the case where $m = 1$, so that we are interested in estimating a single parameter α and make N independent identically distributed observations $x_1, \ldots, x_N$ on α. These x's differ from α by random errors which are assumed to be normal random variables each with mean 0 and known variance, σ^2. Thus the x's are normal random variables, each with mean α and variance σ^2. The problem is to estimate α, the expected value of the x's.

More complex is the case where the x_i's are independent but not identically distributed in that their variances differ. Here x_i is a normal random variable, with mean α and variance σ_i^2, a known number, and again one wishes to estimate α, the expected value of the x's.

Even more general is the case where the x_i's are correlated, so that the vector $(x_1, \ldots, x_N)$ has a multivariate normal distribution with unknown mean vector $\alpha(1, \ldots, 1)$ and known covariance matrix. Again only the scalar α is to be estimated.

Rather than detail the various degrees of generality of the direct measurement case for $m > 1$, let us state the most common case, that of replicated measurements. Here we observe N m-vectors

$(x_{11}, \ldots, x_{1m})$, $(x_{21}, \ldots, x_{2m})$, $\ldots$, $(x_{N1}, \ldots, x_{Nm})$, where the vectors are uncorrelated with each other, and where the i-th vector has a multivariate normal distribution with unknown mean vector $(\alpha_1, \ldots, \alpha_m)$ and known covariance matrix $\sum_i$, i = 1, ..., N.

Even more general is the case where the N successive replicated observations on each of the α_i's are themselves correlated. In this case we form the mN-vector $x = (x_{11}, \ldots, x_{1m}, x_{21}, \ldots, x_{2m}, \ldots, x_{N1}, \ldots, x_{Nm})$ and assume that it has a normal distribution with unknown mean vector $\alpha = (\alpha_1, \ldots, \alpha_m, \alpha_1, \ldots, \alpha_m, \ldots, \alpha_1, \ldots, \alpha_m)$ and a known covariance matrix Ω which is not block diagonal.

We can subsume all these cases in a very general statement of the estimation problem, as follows. Let x be a vector of p observations, with multivariate normal distribution with unknown mean vector α and known* covariance matrix Ω. Our problem is to estimate the vector α, and the method of maximum likelihood requires that we maximize the likelihood function of x, i.e., that we maximize

$$\phi(x) = \frac{|\Omega|^{-\frac{1}{2}}}{(2\pi)^{p/2}} \quad e^{-\frac{1}{2}(x-\alpha)\ \Omega^{-1}(x-\alpha)'}$$

with respect to α. This reduces to the minimization with respect to α of the quadratic form

$$(x-\alpha)\ \Omega^{-1}(x-\alpha)'$$

*In practice, Ω may not be known, but only estimated, hopefully from data independent of x, and this estimate is used as the true value of Ω, rather than as a random variable. Though such usage in the strict mathematical sense invalidates the subsequent analysis, the practical effect of this usage should be negligible if the estimate of Ω is based on a considerable amount of data on instrumental precision.

Writing the general quadratic form to be minimized as $(x-\alpha)\ \Omega^{-1}(x-\alpha)'$, without additional words of caution, is a bit misleading, as it is not clear from the quadratic form whether a p-vector x is composed of one observation on each of p different α's (so that p = m) or N observations on each of p/N different α's (so that p = mN). In the latter case if we first write the vector α as $\alpha = (\alpha_1, \ldots, \alpha_p)$ we have as constraints among these α_i's that

$$\alpha_1 = \alpha_{p/N + 1} = \alpha_{2p/N + 1} = \cdots = \alpha_{(N-1)\ p/N + 1}$$

$$\alpha_2 = \alpha_{p/N + 2} = \alpha_{2p/N + 2} = \cdots = \alpha_{(N-1)\ p/N + 2}$$

etc., whereas in the former case there are no constraints among the α's.

And even when x is a vector of a single observation on each of m different α's, the α_i's need not be functionally independent parameters. For example, we might measure with error the three sides and three angles of a triangle, so that m = 6, yet the six parameters $\alpha_1, \ldots, \alpha_6$ are not functionally independent. Such functional constraints, as well as constraints on the α's of the type described above due to replication, must be taken into account in the minimization of the quadratic form $(x-\alpha)\ \Omega^{-1}(x-\alpha)'$.

We make the effect of redundancy among parameters and replication among our observations upon the minimization problem explicit by introducing constraint equations. We suppose that there are m functionally independent parameters among the p parameters $\alpha_1, \ldots, \alpha_p$ to be

estimated, and that the functional dependence of the remaining $s = p - m$ parameters is given by functions $\psi_1(\alpha_1, \ldots, \alpha_p) = 0$, $\ldots$, $\psi_s(\alpha_1, \ldots, \alpha_p) = 0$. Then the method of maximum likelihood reduces to that of minimizing $(x-\alpha)\,\Omega^{-1}(x-\alpha)'$ with respect to α subject to the constraints $\psi_i(\alpha) = 0$, $i = 1, \ldots, s$.

A few examples will help illustrate the generality of this formulation.

Example 1 (independent replications): Here we observe $\alpha_1, \ldots, \alpha_m$ N times, so that $p = mN$, and $x = (x_{11}, \ldots, x_{1m}, x_{21}, \ldots, x_{2m}, \ldots, x_{N1}, \ldots, x_{Nm})$, and

$$\Omega = \begin{bmatrix} \sum_1 & & 0 \\ & \sum_2 & \\ & & \ddots \\ 0 & & \sum_m \end{bmatrix}$$

where $\sum_i$ is the covariance matrix of $(x_{i1}, \ldots, x_{im})$.

We write α as $\alpha = (\alpha_1, \ldots, \alpha_p)$.

Then

$$(x - \alpha)\,\Omega^{-1}(x - \alpha)' = \sum_{i=1}^{N} \sum_{j,k=1}^{m} \sigma_i^{jk}\,(x_{ij} - \alpha_{m(i-1)+j})(x_{ik} - \alpha_{m(i-1)+k})$$

where σ_i^{jk} is the (j,k)-th element of $\sum_i^{-1}$

In this example $s = m(N-1)$, and the constraints are of the form

$$\begin{aligned}
\psi_1(\alpha) &= \alpha_1 - \alpha_{m+1} = 0 \\
\psi_2(\alpha) &= \alpha_1 - \alpha_{2m+1} = 0 \\
&\vdots \\
\psi_m(\alpha) &= \alpha_1 - \alpha_{(N-1)m+1} = 0 \\
\psi_{m+1}(\alpha) &= \alpha_2 - \alpha_{m+2} = 0 \\
&\vdots \\
\psi_{m(N-1)}(\alpha) &= \alpha_m - \alpha_p = 0
\end{aligned}$$

Of course, one need not go through this formalism for this example, as the constraints can easily be substituted into the quadratic form, leaving us with the unconstrained minimization problem of minimizing

$$\sum_{i=1}^{N} \sum_{j,k=1}^{m} \sigma_i^{jk}(x_{ij} - \alpha_j)(x_{ix} - \alpha_k)$$

with respect to $\alpha_1, \ldots, \alpha_m$.

Example 2 (unequal sample sizes): Here for simplicity we let $m = 2$ and observe one α n_1 times and another α n_2 times, so that $p = n_1 + n_2$ and $x = (x_{11}, \ldots, x_{n_1 1}, x_{12}, \ldots, x_{n_2 2})$. Let Ω be the known $(n_1 + n_2)x(n_1 + n_2)$ covariance matrix of x, and write $\alpha = (\alpha_1, \ldots, \alpha_p)$. Then $s = n_1 + n_2 - 2$ and the constraints are of the form

$$\begin{aligned}
\psi_1(\alpha) &= \alpha_1 - \alpha_2 &&= 0 \\
\psi_2(\alpha) &= \alpha_1 - \alpha_3 &&= 0 \\
&\vdots \\
\psi_{n_1-1}(\alpha) &= \alpha_1 - \alpha_{n_1} &&= 0 \\
\psi_{n_1}(\alpha) &= \alpha_{n_1+1} - \alpha_{n_1+2} &&= 0 \\
&\vdots \\
\psi_{n_1+n_2-2}(\alpha) &= \alpha_{n_1+1} - \alpha_{n_1+n_2} &&= 0
\end{aligned}$$

Again the constraints can be substituted directly into the quadratic form $(x - \alpha)\,\Omega^{-1}(x - \alpha)'$.

When Ω is block-diagonal, i.e., when the observations on one of the parameters are uncorrelated with those on the other, but all observations on a single parameter are correlated, then

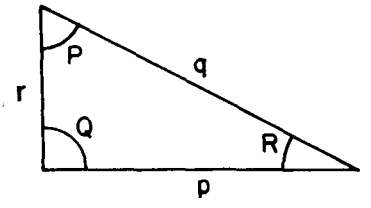
$$\Omega = \begin{bmatrix} \sum_1 & 0 \\ 0 & \sum_2 \end{bmatrix}$$

and the quadratic form reduces to

$$(x - \alpha)\,\Omega^{-1}(x - \alpha)' = \sum_{j,k=1}^{n_1} \sigma_1^{jk}(x_{j1} - \alpha_1)(x_{k1} - \alpha_1) + \sum_{j,k=1}^{n_2} \sigma_2^{jk}(x_{j2} - \alpha_{n_1+1})(x_{k2} - \alpha_{n_1+1}).$$

Example 3 (replication and redundancy): We will describe here a generalization of a surveying problem taken from Deming.[11]

In this example we are surveying a plane triangle



and have 4 observations on P, 2 observations on Q and R, and one observation on p, q, and r. Here p = 11, and we write

$$x = (x_{11}, x_{21}, x_{31}, x_{41}, x_{12}, x_{22}, x_{13}, x_{23}, x_4, x_5, x_6)$$

$$\alpha = (\alpha_1, \ldots, \alpha_{11})$$

and the constraints due to replication are

$$\psi_1(\alpha) = \alpha_1 - \alpha_2 = 0$$
$$\psi_2(\alpha) = \alpha_1 - \alpha_3 = 0$$
$$\psi_3(\alpha) = \alpha_1 - \alpha_4 = 0$$
$$\psi_4(\alpha) = \alpha_5 - \alpha_6 = 0$$
$$\psi_5(\alpha) = \alpha_7 - \alpha_8 = 0$$

In addition, there are three constraints due to the functional dependence of the sides and angles of a triangle:

$$\psi_6(\alpha) = \frac{\sin \alpha_1}{\alpha_9} - \frac{\sin \alpha_5}{\alpha_{10}} = 0$$

$$\psi_7(\alpha) = \frac{\sin \alpha_1}{\alpha_9} - \frac{\sin \alpha_7}{\alpha_{11}} = 0$$

$$\psi_8(\alpha) = \alpha_1 + \alpha_5 + \alpha_7 - \pi = 0$$

The quadratic form $Q = (x-\alpha)\,\Omega^{-1}(x-\alpha)'$ is to be minimized subject to these constraints to obtain the maximum likelihood estimate of α.

The first five constraints are of simple enough form that one can substitute them directly into Q. When Ω is block diagonal of the form

$$\Omega = \begin{bmatrix} \sum_1 & & & & & & \\ & \sum_2 & & & & O & \\ & & \sum_3 & & & & \\ & & & \sigma_4^2 & & & \\ & & & & \sigma_5^2 & & \\ & O & & & & \sigma_6^2 & \end{bmatrix}$$

where $\sum_1$ is a 4x4 matrix, $\sum_2$ and $\sum_3$ are 2x2 matrices, and σ_4^2, σ_5^2 and σ_6^2 are scalars, then Q breaks up into six sums:

$$Q = \sum_{j,k=1}^{4} \sigma_1^{jk}(x_{j1} - \alpha_1)(x_{k1} - \alpha_1) + \sum_{j,k=1}^{2} \sigma_2^{jk}(x_{j2} - \alpha_5)(x_{k2} - \alpha_5)$$

$$+ \sum_{j,k=1}^{2} \sigma_3^{jk}(x_{j3} - \alpha_7)(x_{k2} - \alpha_7) + \frac{(x_9 - \alpha_9)^2}{\sigma_4^2} + \frac{(x_{10} - \alpha_{10})^2}{\sigma_5^2}$$

$$+ \frac{(x_{11} - \alpha_{11})^2}{\sigma_6^2}.$$

When $\sum_i = \sigma_i^2 I$, then the i-th sum reduces further to the sum of squares of deviations of the x_{ji} from their means, divided by σ_i^2.

In the case of indirect measurement, we suppose that, though the x's described above are direct measurements on the vector $\alpha = (\alpha_1, \ldots, \alpha_p)$, these α's are functions of a vector $\mu = (\mu_1, \ldots, \mu_t)$, and our problem is to estimate μ. Since there are p-s functionally

independent α's, in order for these α's to be functions of t functionally independent μ's we must have $t \leq p-s$. When $t > p-s$ the μ's are functionally dependent and this dependence must be taken into account by further constraints on the μ_j's. We shall not consider this case, except to point out that it can easily be handled by extending the procedure we outline to include these constraints.

From our mathematical model we determine the t independent functions $f_i(\alpha, \mu) = 0$ which relate the directly observed parameters α with the indirectly observed parameters μ. Our problem in this case is to minimize $(x - \alpha)\, \Omega^{-1} (x - \alpha)'$ with respect to α and μ subject to the s+t constraints $f_i(\alpha, \mu) = 0$, $i = 1, \ldots, t$, $\psi_j(\alpha) = 0$, $j = 1, \ldots, s$. We shall discuss methods for performing this minimization and then obtaining the maximum likelihood estimates of α (in the direct measurement case) or μ (in the indirect measurement case) in the next section.

Example 4 (indirect measurement): To illustrate the case of indirect measurement we present a modification of Example 3. In this example, we are surveying the same triangle as in Example 3, except that we have only the observations on P, Q, R, and p, and are interested in estimating q and r. Here $p = 9$, $s = 5$, and we write

$$x = (x_{11}, x_{21}, x_{31}, x_{41}, x_{12}, x_{22}, x_{13}, x_{23}, x_9)$$

$$\alpha = (\alpha_1, \ldots, \alpha_9)$$

$$\psi_1(\alpha) = \alpha_1 - \alpha_2 = 0$$

$$\psi_2(\alpha) = \alpha_1 - \alpha_3 = 0$$

$$\psi_3(\alpha) = \alpha_1 - \alpha_4 = 0$$

$$\psi_4(\alpha) = \alpha_5 - \alpha_6 = 0$$

$$\psi_5(\alpha) = \alpha_7 - \alpha_8 = 0$$

$$\psi_5(\alpha) = \alpha_1 + \alpha_5 + \alpha_7 - \pi = 0$$

Our vector μ is composed of the unknowns of interest, q and r, so that t = 2. We write $\mu_1 = q$, $\mu_2 = r$, and

$$f_1(\alpha, \mu) = \frac{\sin \alpha_1}{\alpha_9} - \frac{\sin \alpha_5}{\mu_1} = 0$$

$$f_2(\alpha,\mu) = \frac{\sin \alpha_1}{\alpha_9} - \frac{\sin \alpha_7}{\mu_2} = 0$$

IV. GENERAL LEAST SQUARES ESTIMATION

As seen earlier, the most general estimation problem is one involving indirect measurements, where x is a p-vector of observations on α, μ is a t-vector $(\mu_1, \ldots, \mu_t)$, where $f_i(\alpha, \mu) = 0$, $i = 1, \ldots, t$, and $\psi_j(\alpha) = 0$, $j = 1, \ldots, s$. The problem is to estimate α and μ subject to the s + t constraints given above.

The method of ordinary least squares requires that one minimize the quadratic form $Q = (x - \alpha)(x - \alpha)'$ with respect to α and μ, subject to the s + t constraints $\psi_j(\alpha) = 0$, $f_i(\alpha, \mu) = 0$. More generally, one can assign weights d_i to each of the x_i and require that the quadratic form $Q^* = (x - \alpha)\, D\, (x - \alpha)'$ be minimized, where D is a diagonal matrix of weights d_i. One can be still more general and define the general least squares problem as one of minimizing the quadratic form $(x - \alpha)\, W\, (x - \alpha)'$ with respect to α and μ, where W is a known symmetric positive definite matrix of weights.

Let $\hat{\alpha}(W)$ and $\hat{\mu}(W)$ denote the minimizing α and μ for the weight matrix W. It can be shown that choice of the inverse of the covariance matrix of x as the weight matrix will lead to estimates of α and μ which have asymptotically smallest variance <u>among all estimates of the form</u> $\hat{\alpha}(W)$, $\hat{\mu}(W)$. This is not to say that these estimates have asymptotically smallest variance. As pointed out earlier, it is the maximum likelihood estimate which has this more desired property.

If we call Ω the covariance matrix of x, then the "best" quadratic form to be minimized is

$$(x - \alpha)\, \Omega^{-1} (x - \alpha)'$$

which is identical with the quadratic form to be minimized in maximum likelihood estimation for the multivariate normal distribution. Thus when the distribution of the measurements is multivariate normal, the least squares estimates $\hat{\alpha}(\Omega^{-1})$, $\hat{\mu}(\Omega^{-1})$ have asymptotically smallest variance among all estimates of α and μ, not merely among all estimates of the form $\hat{\alpha}(W)$, $\hat{\mu}(W)$.

Minimization of this quadratic form subject to the constraints $f_j(\alpha, \mu) = 0$, $\psi_i(\alpha) = 0$ is accomplished by the method of Lagrange multipliers, which requires that one find the saddlepoint of the Lagrangian

$$S = (x - \alpha)\, \Omega^{-1} (x - \alpha)' - 2 \sum_{j=1}^{s} \lambda_j \psi_j(\alpha) - 2 \sum_{\ell=1}^{t} \zeta_\ell f_\ell(\alpha, \mu)$$

where $\lambda_1, \ldots, \lambda_s$ and $\zeta_1, \ldots, \zeta_t$ are Lagrange multipliers. Here we equate to zero the derivative of S with respect to both the α_i and the μ_j to obtain the p + t equations

$$\left[(x - \alpha)\, \Omega^{-1}\right]_i = \sum_{j=1}^{s} \lambda_j \frac{\partial \psi_j(\alpha)}{\partial \alpha_i} + \sum_{\ell=1}^{t} \zeta_\ell \frac{\partial f_\ell(\alpha, \mu)}{\partial \alpha_i} \qquad i = 1, \ldots, p$$

$$\sum_{\ell=1}^{t} \zeta_\ell \frac{\partial f_\ell(\alpha, \mu)}{\partial \mu_j} = 0 \qquad j = 1, \ldots, t$$

where $\left[(x - \alpha)\, \Omega^{-1}\right]_i$ is the i-th element of the vector $(x - \alpha)\, \Omega^{-1}$.

In general, these equations can only be solved by numerical methods. However, a judicious use of Newton's method can simplify matters, after rewriting these equations in matrix notation by introducing the vectors $F(\alpha, \mu) = \left(f_1(\alpha, \mu), \ldots, f_t(\alpha, \mu) \right)$, $\psi(\alpha) = \left(\psi_1(\alpha), \ldots, \psi_s(\alpha) \right)$, $\Lambda = (\lambda_1, \ldots, \lambda_s)$ and $Z = (\zeta_1, \ldots, \zeta_t)$, the t x t matrix $F_\mu(\alpha, \mu)$ with $f_{\mu ij} = [\partial f_i(\alpha, \mu)] / [\partial \mu_j]$ the s x p matrix ψ_α with $\psi_{\alpha ij} = [\partial \psi_i(\alpha)] / [\partial \alpha_j]$, and the t x p matrix $F_\alpha(\alpha, \mu)$ with $f_{\alpha ij} = [\partial f_i(\alpha, \mu)] / [\partial \alpha_j]$.

Let α° be an initial guess at the value of the minimizing α, where $\psi_j(\alpha^\circ)$ need not equal zero for $j = 1, \ldots, s$, and let μ° be an initial guess at the value of the minimizing μ. Then we can approximate $\psi_j(\alpha)$ by

$$\psi_j(\alpha) \simeq \psi_j(\alpha^\circ) + \sum_{i=1}^{p} \left. \frac{\partial \psi_j(\alpha)}{\partial \alpha_i} \right|_{\alpha = \alpha^\circ} (\alpha_i - \alpha_i^\circ)$$

and $f_j(\alpha, \mu)$ by

$$f_j(\alpha, \mu) \simeq f_j(\alpha_o, \mu_o) + \sum_{i=1}^{t} \left. \frac{\partial f_j(\alpha, \mu)}{\partial \mu_i} \right|_{\substack{\alpha = \alpha^\circ \\ \mu = \mu^\circ}} (\mu_i - \mu_i^\circ) + \sum_{i=1}^{p} \left. \frac{\partial f_j(\alpha, \mu)}{\partial \alpha_i} \right|_{\substack{\alpha = \alpha^\circ \\ \mu = \mu^\circ}} (\alpha_i - \alpha_i^\circ)$$

Then S is approximated by

$$(x - \alpha)\, \Omega^{-1} (x - \alpha)' - 2\Lambda \left[\psi'(\alpha^\circ) + \psi_\alpha(\alpha_o)(\alpha - \alpha^\circ)' \right]$$
$$-2Z \left[F'(\alpha^\circ, \mu^\circ) + F_\alpha(\alpha^\circ, \mu^\circ)(\alpha - \alpha^\circ)' + F_\mu(\alpha^\circ, \mu^\circ)(\mu - \mu^\circ)' \right]$$

Equating the derivatives of S with respect to α and μ to zero, we obtain the matrix equations

$$(x - \alpha)\,\Omega^{-1} - \Lambda\,\psi_{\alpha}(\alpha^{\circ}) - Z F_{\alpha}(\alpha^{\circ}, \mu^{\circ}) = 0$$

$$Z\,F_{\mu}(\alpha^{\circ}, \mu^{\circ}) = 0$$

As the μ's are functionally independent, the t x t matrix $F_{\mu}(\alpha^{\circ}, \mu^{\circ})$ is non-singular, since it is the Jacobian of the transformation from the space of μ's to the space of f's. Thus $Z = 0$.

To determine Λ we substitute the value

$$\alpha = x - \Lambda\,\psi_{\alpha}(\alpha^{\circ})\,\Omega$$

into the linearized constraint equation for α to obtain

$$\psi(\alpha^{\circ}) + (x - \Lambda\,\psi_{\alpha}(\alpha^{\circ})\,\Omega - \alpha^{\circ})\,\psi_{\alpha}'(\alpha^{\circ}) = 0$$

or

$$\Lambda = \left[\psi(\alpha^{\circ}) + (x - \alpha^{\circ})\,\psi_{\alpha}'(\alpha^{\circ})\right]\left[\psi_{\alpha}(\alpha^{\circ})\,\Omega\,\psi_{\alpha}(\alpha^{\circ})\right]^{-1}$$

provided $\psi_{\alpha}(\alpha^{\circ})\,\Omega\,\psi_{\alpha}'(\alpha^{\circ})$ is non-singular. Then

$$\alpha = x - \left[\psi(\alpha^{\circ}) + (x - \alpha^{\circ})\,\psi_{\alpha}'(\alpha^{\circ})\right]\left[\psi_{\alpha}(\alpha^{\circ})\,\Omega\,\psi_{\alpha}'(\alpha^{\circ})\right]^{-1}\psi_{\alpha}(\alpha^{\circ})\,\Omega$$

$$= x\left[I - \psi_{\alpha}'(\alpha^{\circ})\left[\psi_{\alpha}(\alpha^{\circ})\,\Omega\,\psi_{\alpha}'(\alpha^{\circ})\right]^{-1}\psi_{\alpha}(\alpha^{\circ})\,\Omega\right]$$

$$+\left[\alpha^{\circ}\psi_{\alpha}'(\alpha^{\circ}) - \psi(\alpha_{\circ})\right]\left[\psi_{\alpha}(\alpha^{\circ})\,\Omega\,\psi_{\alpha}'(\alpha^{\circ})\right]^{-1}\psi_{\alpha}(\alpha^{\circ})\,\Omega$$

The vector μ is found by solving the other constraint equation to be

$$\mu = \mu^{\circ} - \left[(\alpha - \alpha^{\circ})\,F_{\alpha}'(\alpha^{\circ}, \mu^{\circ}) + F'(\alpha^{\circ}, \mu^{\circ})\right]\left[F_{\mu}'(\alpha^{\circ}, \mu^{\circ})\right]^{-1}$$

These values of α and μ satisfy the approximate constraints, though not the exact constraints, even if α° and μ° did do so. However, these values of α and μ are used as α° and μ° and the procedure is iterated until it converges on a value $\hat{\alpha}$ of α and $\hat{\mu}$ of μ, which then should both satisfy the exact constraint and minimize $(x - \alpha)\ \Omega^{-1}(x - \alpha)'$.

It is clear from the above that the final estimate $\hat{\alpha}$ of α will be of the form $x\ A + B$, where A is a matrix of the form

$$A = \left[I - \psi_\alpha' \ (\psi_\alpha \Omega \ \psi_\alpha')^{-1} \ \psi_\alpha \ \Omega \right]$$

Since Ω is the covariance matrix of x, we see that the covariance matrix of $\hat{\alpha}$ is $A' \ \Omega \ A = \Omega - \Omega \ \psi_\alpha' \ (\psi_\alpha \ \Omega \ \psi_\alpha)^{-1} \ \psi_\alpha \ \Omega = \Omega \ A$, an easily computed quantity.

The vector $\hat{\mu}$ is of the form $\hat{\alpha}\ C + D$, as seen from the above, where

$$C = - \left[F_\mu^{-1}(\alpha^\circ, \mu^\circ) \ F_\alpha(\alpha^\circ, \mu^\circ) \right]'$$

so that the covariance matrix of $\hat{\mu}$ is

$$C' \ A' \ \Omega \ A \ C = C' \ \Omega \ A \ C$$

The cross-covariance matrix between $\hat{\alpha}$ and $\hat{\mu}$ is given by $A' \ \Omega \ A \ C = \Omega \ A \ C$, so that the covariance matrix of $(\hat{\alpha}, \hat{\mu})$ is

$$\begin{bmatrix} \Omega \ A & \Omega \ A \ C \\ C'A' \ \Omega & C' \ \Omega \ A \ C \end{bmatrix}$$

V. THE PROBLEM OF WEIGHTING--COMBINING RESULTS FROM SEVERAL EXPERIMENTS

In least squares estimation, the covariance matrix associated with the observations is important in giving the proper weight to the different observations. It is clear from the form of the estimate of α that Ω need be known only up to a scale factor, so that only relative values of the variances and covariances are important. The covariance matrix of the estimate will then also be known only up to a scale factor. However, as we shall see, if the covariance matrix accompanying the estimates from one experiment is to provide the basis for comparing and making an overall adjustment among several experiments, then relative values of the variances and covariances will not suffice; their absolute level must also be known.

Suppose we have additional observations on α's, either as replications of observations on α's already members of the vector $\alpha = (\alpha_1, \ldots, \alpha_p)$ or as observations on a new α, say α_{p+1}, which may even be a function of a new μ, say μ_{t+1}, with an associated constraint $f_{t+1}(\alpha, \mu) = 0$. One sees from our formulation above that these can easily be handled by enlarging the x vector, α vector, and covariance matrix Ω appropriately. In most cases this is the only way to handle the problem of combining data from many experiments.

In the special case where the N successive experiments are independent and replicates of each other and the covariance matrices for each experiment are diagonal, however, one can easily check that the problem reduces to one in which one has observed a single p vector $x = (\bar{x}_1, \ldots, \bar{x}_p)$, where

$$\bar{x}_i = \frac{\sum_{j=1}^{N} \sigma_j^i \, x_{ij}}{\sum_{j=1}^{N} \sigma_j^i} \quad , \; i = 1, \ldots, p$$

and σ_j^i is the reciprocal of the variance of the j-th observation x_{ij} on α_i, so that x_i is a weighted average of the observations on α_i.

To illustrate how to combine two independent experiments, we will assume that in one experiment two parameters have been estimated, say $\hat{\alpha}_1$, $\hat{\alpha}_2$ with a covariance matrix Σ_1. In another experiment, two parameters $\hat{\alpha}_3$, $\hat{\alpha}_4$ have been estimated with covariance matrix Σ_2. In this example, we will assume that $\hat{\alpha}_2$ and $\hat{\alpha}_3$ are estimates of the same parameter.

In this case, we take $x = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4)$, $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, and $\psi_1(\alpha) = \alpha_2 - \alpha_3$. The covariance matrix of x is

$$\Omega = \begin{bmatrix} \sigma_{11}^{(1)} & \sigma_{12}^{(1)} & 0 & 0 \\ \sigma_{21}^{(1)} & \sigma_{22}^{(1)} & 0 & 0 \\ 0 & 0 & \sigma_{11}^{(2)} & \sigma_{12}^{(2)} \\ 0 & 0 & \sigma_{21}^{(2)} & \sigma_{22}^{(2)} \end{bmatrix}$$

Applying the general least squares method, we would then have

$$\psi_\alpha = (0, 1, -1, 0)$$

Our estimate of α is given by

$$\tilde{\alpha} = x - \Lambda \, \psi_\alpha \, \Omega$$

and Λ is determined by solving the constraint equation $\psi(\tilde{\alpha}) = \psi(x - \Lambda \psi_{\alpha} \Omega) = 0$ for the unknown Λ. In this case

$$\psi(\tilde{\alpha}) = \hat{\alpha}_2 - \hat{\alpha}_3 - \lambda(\sigma_{22}^{(1)} + \sigma_{11}^{(2)}) = 0$$

so that

$$\lambda = \frac{\hat{\alpha}_2 - \hat{\alpha}_3}{\sigma_{22}^{(1)} + \sigma_{11}^{(2)}}$$

and $$\tilde{\alpha} = x - \frac{\hat{\alpha}_2 - \hat{\alpha}_3}{\sigma_{22}^{(1)} + \sigma_{11}^{(2)}} \quad (\sigma_{21}^{(1)}, \sigma_{22}^{(1)}, -\sigma_{11}^{(2)}, -\sigma_{12}^{(2)})$$

is the combined least squares estimate of α.

It is obvious that $\tilde{\alpha}_2 = \tilde{\alpha}_3$, and for consistency we also have new estimates for α_1 and α_4. This simple example illustrates how one may use variances and covariances in weighting the results from more than one experiment to obtain new over-all estimates for the separately determined quantities.

Accompanying this new estimate will be a new covariance matrix, determined as follows: write $\tilde{\alpha}$ in matrix form as

$$\tilde{\alpha} = \frac{x}{\sigma_{22}^{(1)} + \sigma_{11}^{(2)}} \begin{bmatrix} \sigma_{22}^{(1)} + \sigma_{11}^{(2)} & 0 & 0 & 0 \\ -\sigma_{21}^{(1)} & \sigma_{11}^{(2)} & \sigma_{11}^{(2)} & \sigma_{12}^{(2)} \\ \sigma_{21}^{(1)} & \sigma_{22}^{(1)} & \sigma_{22}^{(1)} & -\sigma_{12}^{(2)} \\ 0 & 0 & 0 & \sigma_{22}^{(1)} + \sigma_{11}^{(2)} \end{bmatrix}$$

$$= x\, A,$$

say. Then the covariance matrix of $\tilde{\alpha}$ is $A' \Omega A$.

The variance of $\tilde{\alpha}_2 (= \tilde{\alpha}_3)$ is

$$\frac{\sigma_{22}^{(1)} \sigma_{11}^{(2)}}{\sigma_{22}^{(1)} + \sigma_{11}^{(2)}}$$

which is smaller than either $\sigma_{22}^{(1)}$ or $\sigma_{11}^{(2)}$, the variances of $\hat{\alpha}_2$ and $\hat{\alpha}_3$, respectively. The variance of $\tilde{\alpha}_1$ is

$$\sigma_{11}^{(1)} - \frac{\left[\sigma_{12}^{(1)}\right]^2}{\sigma_{22}^{(1)} + \sigma_{11}^{(2)}}$$

and the variance of $\tilde{\alpha}_4$ is

$$\sigma_{22}^{(2)} - \frac{\left[\sigma_{21}^{(2)}\right]^2}{\sigma_{22}^{(1)} + \sigma_{11}^{(12)}}$$

thus reducing the variance of the estimates of α_1 and α_4 from those based on the single experiments.

The same results will be achieved in this case by recognizing from the beginning the identity of α_2 and α_3 and dealing directly with the inverses of the covariance matrices Σ_1 and Σ_2. The quadratic form to be minimized is

$$Q = (x - \alpha^*) \, \Omega^{-1} \, (x - \alpha^*)' \, ,$$

where $\alpha^* = (\alpha_1, \alpha_2, \alpha_2, \alpha_4)$, subject to no constraints on the α's. Since

$$\Omega^{-1} = \begin{bmatrix} \Sigma_1^{-1} & 0 \\ 0 & \Sigma_2^{-1} \end{bmatrix}$$

Q can be rewritten as follows: Define new 3 x 3 matrices

$$S_1 = \begin{bmatrix} \Sigma_1^{-1} & & 0 \\ & & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & & \\ 0 & \Sigma_2^{-1} & \end{bmatrix}$$

and vectors $x_1^* = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_4)$, $x_2^* = (\hat{\alpha}_1, \hat{\alpha}_3, \hat{\alpha}_4)$, and $\alpha = (\alpha_1, \alpha_2, \alpha_4)$. Then

$$Q = (x_1^* - \alpha)\ S_1(x_1^* - \alpha)' + (x_2^* - \alpha)\ S_2(x_2^* - \alpha)'$$

Now setting the derivative of Q with respect to α to zero gives us the equation

$$(x_1^* - \alpha)\ S_1 + (x_2^* - \alpha)\ S_2 = 0$$

or

$$\tilde{\alpha} = (x_1^*\ S_1 + x_2^*\ S_2)(S_1 + S_2)^{-1}$$

The covariance matrix of $\tilde{\alpha}$ is easily determined by first noting that

$$x_1^*\ S_1 + x_2^*\ S_2 = (\hat{\alpha}_1, \hat{\alpha}_2, 0)\ S_1 + (0, \hat{\alpha}_3, \hat{\alpha}_4)\ S_2$$

where the vectors $(\hat{\alpha}_1, \hat{\alpha}_2, 0)$ and $(0, \hat{\alpha}_3, \hat{\alpha}_4)$ are independent.

From this we see that the covariance matrix of $\tilde{\alpha}$ is

$$(S_1 + S_2)^{-1} \left\{ S_1 \begin{bmatrix} \Sigma_1 & & 0 \\ & & 0 \\ 0 & 0 & 0 \end{bmatrix} S_1 + S_2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & & \\ 0 & & \Sigma_2 \end{bmatrix} S_2 \right\} (S_1 + S_2)^{-1}$$

$$= (S_1 + S_2)^{-1}$$

The final results obtained by this procedure will agree in all respects with those obtained previously. It is obvious that either procedure can be extended to cases involving any number of parameters in each experiment and many overlapping parameters. Numerical examples of this method of weighting the results of two or more experiments are given in Appendix II.

VI. IMPLICATIONS AND PROBLEMS IN APPLYING LEAST SQUARES ANALYSIS

When all requirements of a least squares analysis are met, an excellent case can be made for its use in estimating solar system constants. When the requirements are not completely fulfilled, least squares analysis still provides a method of estimation which is at least more precise and reproducible than some methods which have been used. The method makes the estimates satisfy the theoretical constraint equations of the experiments.

Ideally, we would like to require that:

1. All observations represent a sample from a multivariate normal distribution.
2. Systematic errors are small enough, or the results of least squares analysis are insensitive enough, that estimates are not appreciably affected by them.
3. Parameter estimates and the covariance matrices accompanying estimates are the result of proper assessment of errors in the original observations.

However, regardless of the distribution of measurement errors, the method of general least squares outlined and demonstrated in examples here has much to commend its use. With regard to requirement 1, experience shows that the distribution of measurement errors usually tends to closely approximate the normal or Gaussian distribution.

The problem of learning about and doing something about systematic errors is a difficult one, although not always impossible. Care and discretion in the handling of instruments and closer coordination and better communication between data collectors and data processors will often avoid introduction of certain systematic errors. If the presence of systematic errors in measurements is suspected from the beginning, and their functional relationship with other system parameters is known with any degree of certainty, parameters characterizing systematic error may be incorporated into the mathematical model for the system. This essentially means that there are more parameters to be estimated and that there are additional condition equations in the least squares formulation. When there are sufficient data, such an amplification is feasible. However, the introduction of additional undetermined parameters into the model is not always justified when these parameters are already known with a requisite degree of certainty.[12] We cannot give a definite answer to the question of what to do about combining estimates from two (or more) experiments when there is an obvious incompatibility, as evidenced, say, by estimates for the same parameter that differ by more than two or three times the standard deviation as estimated from the most variable experiment. When this situation occurs, as it has in solar systems constant estimation, the tendency is to suspect the presence of systematic errors, but this may not always be the reason for the discrepancy. It is obvious from the analysis of general least squares procedures that incorrect a priori assessment of even <u>relative</u> observation errors can result in incorrect estimates. Also, one may employ the

wrong mathematical model or derive from an adequate mathematical model an inadequate set of condition equations. However, it is possible to incorporate new parameters into a least squares estimate, thus modifying the model to take account of some systematic errors. As measurement technology in space exploration improves, it is almost certain that we will be obliged to modify and in general elaborate on our models in order to incorporate different and better and even just more observations. This can and should be done in a manner which does not require any apology for having used earlier a now discarded model that really served its purpose.

Certain statistical significance tests may be used which, on a probability basis, can provide a means of deciding whether the results of a given experiment should be used in improving estimates of solar system constants. For example, one can use a chi-square test to decide whether two sets of data are from the same universe. Or one can test a more specific hypothesis, that the two sets of data have common error variance. Unfortunately, when such a test suggests rejection of the data there still may be no way to pinpoint the difficulty, and successful employment of the test may require more data than exist.

For the benefit of those concerned with improving estimates, it is important that the experimenter report not only the final estimated values for all parameters estimated, but the complete covariance matrix associated with the estimates. For further evaluation, there should be available also the list of values for all constants assumed to be known and the theoretical relations employed as constraints, or in other words, comprising the mathematical model for the experiment.

When values for constants are adopted for common usage by some group, having good estimates or "best" estimates by some statistical measure is an important, but not the only consideration. Strict adherence to a theoretical relationship may be more important. To adopt new constants, even if better from the standpoint of statistical estimation, may make vast quantities of data less interpretable, and hence less useful. These and other reasons cause responsible national and international bodies to move slowly in adopting new values. We would expect, however, that the increasing information flow from space experiments would not only provide the data, but also set up the requirements for more frequent updating of values for constants. The availability of high-speed computing machinery removes, to some extent at least, the onus of providing new tables based upon newly adopted constants.

Appendix I

EXAMPLES OF GENERAL LEAST SQUARES ESTIMATES

Example 1--Estimates of Directly Observed Parameters

We will use a surveying problem taken from Deming[11] and described as Example 3 of Section III and apply the procedures outlined in Section IV.

In this example, $\Sigma_i = \sigma_i^2 I$, so that x_{11}, x_{21}, x_{31}, and x_{41} can be replaced by $\bar{x}_1 = \sum_{i=1}^{4} x_{1i}/4$, x_{12} and x_{22} can be replaced by $\bar{x}_2 = (x_{12} + x_{22})/2$, and x_{13} and x_{23} can be replaced by $\bar{x}_3 = (x_{13} + x_{23})/2$ in the quadratic form to be minimized. The average of the measured values are:

on P, $\bar{x}_1 = 51^\circ 06' 25'' = .89193597$ rad.

on Q, $\bar{x}_2 = 95^\circ 04' 5'' = 1.65937179$ rad.

on R, $\bar{x}_3 = 33^\circ 49' 5'' = 0.59035731$ rad.

on p, $x_4 = 1723.7$ ft.

on q, $x_5 = 2205.4$ ft.

on r, $x_6 = 1232.7$ ft.

The standard error for measurements are 1 min. of arc. or 00029 rad. for angles and 1 ft. in 10,000 ft. for lengths and proportional to the square root of the distance. Thus the covariance matrix of the vector $(\bar{x}_1, \bar{x}_2, \bar{x}_3, x_4, x_5, x_6)$ is

$$\Omega = \begin{bmatrix} \left(\frac{.00029}{4}\right)^2 & & & & & \\ & \left(\frac{.00029}{2}\right)^2 & & & & \\ & & \left(\frac{.00029}{2}\right)^2 & & & \\ & & & .1724 & & \\ & & & & .2205 & \\ & & & & & .1233 \end{bmatrix}$$

We have measurements on 6 parameters and a set of 3 parameters are necessary to completely determine the triangle, so there are 3 redundant parameters and 3 condition equations. Identifying α with P, α_2 with Q, α_3 with R, α_4 with p, α_5 with q, and α_6 with r, unlike our notation of example 3, Section III, we rename the 3 condition equations as

$$\psi_1(\alpha) = \frac{\sin \alpha_1}{\alpha_4} - \frac{\sin \alpha_2}{\alpha_5} = 0$$

$$\psi_2(\alpha) = \frac{\sin \alpha_1}{\alpha_4} - \frac{\sin \alpha_3}{\alpha_6} = 0$$

$$\psi_3(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 - \pi = 0$$

Then

$$\psi_\alpha(\alpha) = \begin{bmatrix} \frac{\cos \alpha_1}{\alpha_4} & \frac{-\cos \alpha_2}{\alpha_5} & 0 & \frac{-\sin \alpha_1}{\alpha_4^2} & \frac{\sin \alpha_2}{\alpha_5^2} & 0 \\ \frac{\cos \alpha_1}{\alpha_4} & 0 & \frac{-\cos \alpha_3}{\alpha_4} & \frac{-\sin \alpha_1}{\alpha_4^2} & 0 & \frac{\sin \alpha_3}{\alpha_6^2} \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Taking $\alpha^{o} = x$, we obtain

$$\Lambda = \psi(x) \quad \psi_{\alpha}(x)\ \Omega\ \psi'_{\alpha}(x) \quad ^{-1}$$

$$= 7.2674 \times 10^{6} - 1.7321 \times 10^{6} - 1.7079 \times 10^{3}$$

Then $\overset{\wedge}{\alpha}^{(1)} = x - \Lambda\ \psi_{\alpha}(x)\ \Omega$ or

$$\overset{\wedge}{\alpha}' = \begin{bmatrix} .89194 \\ 1.65932 \\ .5903 \\ 1.72345 \times 10^{3} \\ 2.20572 \times 10^{3} \\ 1.23262 \times 10^{3} \end{bmatrix}$$

These represent estimates that are best in the sense that while satisfying the geometrical conditions expressed by the condition equations, they also result in a minimization of the weighted sums of squares of the residuals. It should be noted that these results as well as those in the following examples have not been refined by interation. Since these estimates are functions of the observations, the covariance matrix of the adjusted set is obtainable in terms of quantities that have in general already been computed.

The covariance matrix of $\overset{\wedge}{\alpha}_{1}$, $\Sigma_{\overset{\wedge}{\alpha}}$, is given by

$$\Sigma_{\overset{\wedge}{\alpha}} = \Omega - \Omega\ \psi'_{\alpha}(x)\left[\psi_{\alpha}(x)\ \Omega\ \psi'_{\alpha}(x)\right]^{-1}\ \psi_{\alpha}(x)\ \Omega,$$

or

$$\underset{\alpha}{\Sigma_{\wedge}} = \begin{bmatrix} 1.360^{-8} & -9.959^{-9} & -3.638^{-9} & 1.418^{-5} & -4.102^{-6} & -1.008^{-5} \\ & 2.193^{-8} & -1.197^{-8} & -4.339^{-7} & 1.286^{-5} & -1.243^{-5} \\ & & 1.561^{-8} & -1.374^{-5} & -8.762^{-6} & 2.251^{-5} \\ & & & 7.912^{-2} & 7.607^{-2} & 1.720^{-2} \\ & & & & 1.021^{-1} & 4.236^{-2} \\ & & \text{(Exponents mean powers of 10.)} & & & 6.373^{-2} \end{bmatrix}$$

If the results of an experiment in which redundant data are handled by the general least squares method is presented, this covariance matrix should be appended to give a useful estimate of the "goodness" of the estimates. It provides further a means of relative weighting of results from similar independent experiments or even different experiments when an overlapping number of quantities have been estimated.

Example 2 --Estimates of Indirectly Observed Parameters

The same surveying problem may be used in a modified form to illustrate a case of least squares estimation of indirectly observed parameters. For instance, suppose measurements are made on P, Q, R, and p, leaving q and r as parameters indirectly determined, as described in Example 4, Section III. For convenience, we will use the same data as in Example 1. We will now have Ω as the 4 x 4 diagonal matrix

$$\Omega = \begin{bmatrix} 2.1025 \times 10^{-8} & 0 & 0 & 0 \\ 0 & 4.1050 \times 10^{-8} & & 0 \\ 0 & 0 & 4.1050 \times 10^{-8} & 0 \\ 0 & 0 & 0 & 1.724 \times 10^{-1} \end{bmatrix}$$

Identifying α_1 with P, α_2 with Q, α_3 with R, α_4 with p, μ_1 with q, and μ_2 with r, unlike our notation of Example 4, Section III, we rename the 3 condition equations as

$$\psi_1(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 - \pi = 0$$

$$f_1(\alpha, \mu) = \frac{\sin \alpha_1}{\alpha_4} - \frac{\sin \alpha_2}{\mu_1} = 0$$

$$f_2(\alpha, \mu) = \frac{\sin \alpha_1}{\alpha_4} - \frac{\sin \alpha_3}{\mu_2} = 0$$

Taking $\alpha^o = x$ and $\mu^o = (x_5, x_6)$ of the previous example, we obtain

$$\hat{\alpha}' = \begin{bmatrix} .89192 \\ 1.65934 \\ .59038 \\ 1.7237 \times 10^3 \end{bmatrix}$$

$$\hat{\mu}' = \begin{bmatrix} 2.20608 \times 10^3 \\ 1.23281 \times 10^3 \end{bmatrix}$$

with

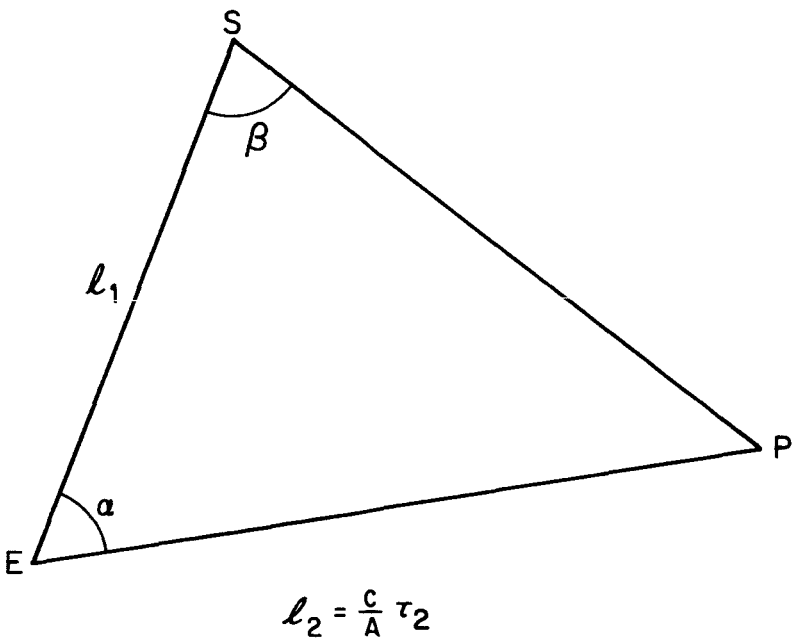
$$\underset{\alpha}{\Sigma}_{\wedge} = \begin{bmatrix} 1.6738 \times 10^{-8} & -8.3692 \times 10^{-9} & -8.3692 \times 10^{-9} & 4.8229 \times 10^{-13} \\ & 2.4710 \times 10^{-8} & -1.6340 \times 10^{-8} & -3.9661 \times 10^{-13} \\ & & 2.4710 \times 10^{-8} & -1.7846 \times 10^{-12} \\ & & & 1.7240 \times 10^{-1} \end{bmatrix}$$

and

$$\underset{\mu}{\Sigma}_{\wedge} = \begin{bmatrix} .33013 & .21893 \\ .21893 & .21895 \end{bmatrix}$$

Example 3--An Estimate of Variance For A Proposed Space Experiment[13]

The experiment proposed originally by A. G. Wilson, and configured and arranged more definitively by Dean Jamison and George Kocher, is represented in elementary form by the diagram below:



An instrumented probe, P, is placed in orbit about the sun S. At various times observations are made on the following parameters:

α = the angle subtended by an earth-sun line and an earth-probe line.

β = the angle subtended by the earth-sun line and a probe sun line.

ℓ_1= the distance from earth to sun in astronomical units. This is not directly observed from the space experiment but is available from tables of the earth's ephemeris and a variance can be assigned to the error in ℓ_1.

τ_2 = the one-way travel time for electromagnetic waves between the earth (E) and probe (P). The purpose of this experiment is to estimate the parameter A/c, the ratio of the astronomical unit of distance in kilometers to the velocity of light in kilometer/second.

The parameters directly measured in this experiment are

$$\alpha = (\alpha,\ \beta,\ \ell_1,\ \tau_2) = (\alpha_1\ \alpha_2,\ \alpha_3\ \alpha_4)$$

with one indirectly determined parameter

$$\mu = A/c$$

The observations are:

$$x^i = (x_1^i,\ x_2^i,\ x_3^i,\ x_4^i)$$

corresponding to

$$\alpha^i = (\alpha_1^i,\ \alpha_2^i,\ \alpha_3^i,\ \alpha_4^i)$$

for the i-th time the set of observations are made. If the error variances on the x^i are $\sigma_1^2,\ \sigma_2^2,\ \sigma_3^2,\ \sigma_4^2$ for all i = 1, ..., n, and observation errors are not correlated, the covariance matrix Ω is a 4n x 4n diagonal matrix,

$$\Omega = \begin{bmatrix} \sigma_1^2 & & & & & & & \\ & \sigma_2^2 & & & & & & \\ & & \sigma_3^2 & & & & & \\ & & & \sigma_4^2 & & & & \\ & & & & \sigma_1^2 & & & \\ & & & & & \sigma_2^2 & & \\ & & & & & & \ddots & \\ & & & & & & & \sigma_4^2 \\ & & & & & & & & \ddots \end{bmatrix}$$

To make the example specific, we will assume n = 3. The covariance matrix Ω will then be a 12 x 12 diagonal matrix.

We rename our parameters, for notational convenience, as

$$\alpha_1 = \alpha_1^1 \qquad \alpha_5 = \alpha_1^2 \qquad \alpha_9 = \alpha_1^3$$

$$\alpha_2 = \alpha_2^1 \qquad \alpha_6 = \alpha_2^2 \qquad \alpha_{10} = \alpha_2^3$$

$$\alpha_3 = \alpha_3^1 \qquad \alpha_7 = \alpha_3^2 \qquad \alpha_{11} = \alpha_3^3$$

$$\alpha_4 = \alpha_4^1 \qquad \alpha_8 = \alpha_4^2 \qquad \alpha_{12} = \alpha_4^3$$

From the geometry and physics of the problem we determine the constraints ψ and f to be

$$\psi_1(\alpha) = \left[\frac{\alpha_4 \sin(\alpha_1 + \alpha_2)}{\alpha_3 \sin \alpha_2}\right]\frac{\sin \alpha_6}{\alpha_8} - \frac{\sin(\alpha_5 + \alpha_6)}{\alpha_7} = 0$$

$$\psi_2(\alpha) = \left[\frac{\alpha_4 \sin(\alpha_1 + \alpha_2)}{\alpha_3 \sin \alpha_2}\right]\frac{\sin \alpha_{10}}{\alpha_{12}} - \frac{\sin(\alpha_9 + \alpha_{10})}{\alpha_{11}} = 0$$

$$f_1(\alpha, \mu) = \left[\frac{\mu \sin \alpha_2}{\alpha_4}\right] - \frac{\sin(\alpha_1 + \alpha_2)}{\alpha_3} = 0$$

The matrix ψ_α will have elements

$$\psi_{\alpha 11} = \frac{\alpha_4 \sin \alpha_6}{\alpha_3 \alpha_8 \sin \alpha_2} \cos(\alpha_1 + \alpha_2)$$

$$\psi_{\alpha 12} = \frac{\alpha_4 \sin \alpha_6 \left[\sin \alpha_2 \cos(\alpha_1 + \alpha_2) - \sin(\alpha_1 + \alpha_2) \cos \alpha_2\right]}{\alpha_3 \alpha_8 \sin \alpha_2}$$

$$\psi_{\alpha 13} = -\frac{\alpha_4 \sin(\alpha_1 + \alpha_2) \sin \alpha_6}{\alpha_3^2 \alpha_8 \sin \alpha_2}$$

$$\psi_{\alpha 14} = \frac{\sin(\alpha_1 + \alpha_2)\ \sin \alpha_6}{\alpha_3\ \alpha_8\ \sin \alpha_2}$$

$$\psi_{\alpha 15} = -\frac{\cos(\alpha_5 + \alpha_6)}{\alpha_7}$$

$$\psi_{\alpha 16} = \frac{\alpha_4\ \sin(\alpha_1 + \alpha_2)}{\alpha_3\ \alpha_8\ \sin\ \alpha_2}\ \cos\ \alpha_6 - \frac{\cos\ (\alpha_5 + \alpha_6)}{\alpha_7}$$

$$\psi_{\alpha 17} = \frac{\sin(\alpha_5 + \alpha_6)}{\alpha_7^2}$$

$$\psi_{\alpha 18} = -\frac{\alpha_4\ \sin(\alpha_1 + \alpha_2)\ \sin\ \alpha_6}{\alpha_8^2\ \alpha_3\ \sin\ \alpha_2}$$

$$\psi_{\alpha 19} = \psi_{\alpha 1,10} = \psi_{\alpha 1,11} = \psi_{\alpha 1,12} = 0$$

$$\psi_{\alpha 21} = \frac{\alpha_4\ \sin\ \alpha_{10}\ \cos(\alpha_1 + \alpha_2)}{\alpha_3\ \alpha_{12}\ \sin\ \alpha_2}$$

$$\psi_{\alpha 22} = \frac{\alpha_4\ \sin\ \alpha_{10}\ \left[\sin\ \alpha_2\ \cos(\alpha_1 + \alpha_2) - \sin(\alpha_1 + \alpha_2)\ \cos\ \alpha_2\right]}{\alpha_3\ \alpha_{12}\ \sin^2\ \alpha_2}$$

$$\psi_{\alpha 23} = \frac{\alpha_4\ \sin(\alpha_1 + \alpha_2)\ \sin\ \alpha_{10}}{\alpha_{12}\ \alpha_3^2\ \sin\ \alpha_2}$$

$$\psi_{\alpha 24} = \frac{\sin(\alpha_1 + \alpha_2)\ \sin\ \alpha_{10}}{\alpha_3\ \alpha_{12}\ \sin\ \alpha_2}$$

$$\psi_{\alpha 25} = \psi_{\alpha 26} = \psi_{\alpha 27} = \psi_{\alpha 28} = 0$$

$$\psi_{\alpha 29} = -\frac{\cos(\alpha_9 + \alpha_{10})}{\alpha_{11}}$$

$$\psi_{\alpha 2,10} = \frac{\alpha_4 \sin(\alpha_1 + \alpha_2)}{\alpha_3 \alpha_{12} \sin \alpha_2} \cos \alpha_{10} - \frac{\cos(\alpha_9 + \alpha_{10})}{\alpha_{11}}$$

$$\psi_{\alpha 2,11} = \frac{\sin(\alpha_9 + \alpha_{10})}{\alpha_{11}^2}$$

$$\psi_{\alpha 2,12} = -\frac{\alpha_4 \sin(\alpha_1 + \alpha_2) \sin \alpha_{10}}{\alpha_{12}^2 \alpha_3 \sin \alpha_2}$$

The F_α matrix will have the elements

$$f_{\alpha 11} = -\frac{\cos(\alpha_1 + \alpha_2)}{\alpha_3}$$

$$f_{\alpha 12} = \frac{\mu}{\alpha_4} \cos \alpha_2 - \frac{\cos(\alpha_1 + \alpha_2)}{\alpha_3}$$

$$f_{\alpha 13} = \frac{\sin(\alpha_1 + \alpha_2)}{\alpha_3^2}$$

$$f_{\alpha 14} = -\frac{\mu \sin \alpha_2}{\alpha_4^2}$$

$$f_{\alpha 15} = f_{\alpha 1,12} = 0$$

and

$$f_\mu = f_{\mu 1} = \frac{\sin \alpha_2}{\alpha_4}$$

Using the results of Section IV, we see that to determine the variance of $\hat{\mu}$ we must compute the matrices

$$A = I - \psi_\alpha' \left[\psi_\alpha \Omega \psi_\alpha' \right]^{-1} \psi_\alpha \Omega$$

and

$$C = - \left[F_\mu^{-1} F_\alpha \right]'$$

Then the 1 x 1 covariance matrix of the estimate of μ is given by

$$\sigma^2_{\hat{\mu}} = C' \Omega A C$$

The standard deviations of the measurements are:

1 sec. of arc in measurements of α

5 sec. of arc in measurements of β

10^{-7} a.u. in measurements of ℓ_1

10^{-6} sec. in measurements of τ_2

Based on these assumptions and the following observations,

$\alpha_1 = 29^\circ.8$	$\alpha_5 = 33^\circ.6$	$\alpha_9 = 65^\circ.8$
$\alpha_2 = 24^\circ.7$	$\alpha_6 = 55^\circ.9$	$\alpha_{10} = 33^\circ.3$
$\alpha_3 = 1.0$	$\alpha_7 = 1.0$	$\alpha_{11} = 1.0$
$\alpha_4 = 256.12531$	$\alpha_8 = 413.21784$	$\alpha_{12} = 277.45447$

we find that the variance of $\hat{\mu}$ is $.4787 \times 10^{-4}$ sec.^{2}/a.u.^{2}

If we assume that c, the velocity of light, is a known constant c = 299792.5 km/sec., we find that the standard deviation of our estimate of the a.u. is 2074 km, using just the three sets of measurements.

Appendix II

EXAMPLE OF WEIGHTING THE RESULTS OF TWO EXPERIMENTS

For the purpose of illustration, we will assume that the results of Examples 1 and 2 of Appendix I represent the analysis of two independent experiments. The fact that some numbers are the same for the two sets of measurements will not make any difference for this purpose. The results of the first example are the estimates of P,Q,R,p,q,r and their accompanying covariance matrices. These are, respectively, the vector $\hat{\underset{\sim}{\alpha}}$ and the matrix $\Sigma_{\hat{\underset{\sim}{\alpha}}}$. In the second example, the results are other estimates for P,Q,R,p, also represented by a vector $\hat{\underset{\sim}{\alpha}}$ and estimates for q,r, represented by the vector $\hat{\underset{\sim}{\mu}}$. The corresponding covariance matrices are $\Sigma_{\hat{\underset{\sim}{\alpha}}}$ and $\Sigma_{\hat{\underset{\sim}{\mu}}}$.

For the purpose of using and weighting the results of both examples to determine a new estimate, we construct a new 12 x 12 covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma^{(1)}_{\hat{\underset{\sim}{\alpha}}} & & 0 \\ & \Sigma^{(2)}_{\hat{\underset{\sim}{\alpha}}} & \\ 0 & & \Sigma_{\hat{\underset{\sim}{\mu}}} \end{bmatrix}$$

where $\Sigma^{(1)}_{\hat{\underset{\sim}{\alpha}}}$ is the 6 x 6 covariance matrix $\Sigma_{\hat{\underset{\sim}{\alpha}}}$ for Example 1, and $\Sigma^{(2)}_{\hat{\underset{\sim}{\alpha}}}$ is the 4 x 4 covariance matrix $\Sigma_{\hat{\underset{\sim}{\alpha}}}$ from Example 2. The zero cross-covariance between $\hat{\underset{\sim}{\alpha}}$ and $\hat{\underset{\sim}{\mu}}$ of Example 2 represents a slight approximation. The detailed procedure from here is the same as Example 1 in Appendix I with the condition equations now simply

$$\psi_1(\alpha) = \alpha_1 - \alpha_7 = 0$$

$$\psi_2(\alpha) = \alpha_2 - \alpha_8 = 0$$

$$\vdots \qquad \vdots$$

$$\psi_6(\alpha) = \alpha_6 - \alpha_{12} = 0$$

$\psi_\alpha(\alpha)$ is thus a 6 x 12 matrix whose elements are 1's, -1's and zeros. Omitting the detailed intermediate results, we have for the new estimates,

$$\hat{\alpha}_1 = .89194403$$

$$\hat{\alpha}_2 = 1.65931298$$

$$\hat{\alpha}_3 = .59033445$$

$$\hat{\alpha}_4 = 1.723444 \times 10^3$$

$$\hat{\alpha}_5 = 2.205719 \times 10^3$$

$$\hat{\alpha}_6 = 1.232612 \times 10^3$$

and the following covariance matrix,

$$\Sigma_{\hat{\alpha}} = \begin{bmatrix} 6.887^{-9} & -5.187^{-9} & -1.700^{-9} & 5.711^{-6} & -3.930^{-6} & -5.893^{-6} \\ & 1.009^{-8} & -4.906^{-9} & -2.235^{-6} & 4.392^{-6} & -5.465^{-6} \\ & & 6.606^{-9} & -3.477^{-6} & -4.622^{-7} & 1.136^{-5} \\ & & & 4.068^{-2} & 4.232^{-2} & 1.702^{-2} \\ & & & & 6.025^{-2} & 3.332^{-2} \\ & & & & & 3.892^{-2} \end{bmatrix}$$

(Exponents mean powers of 10)

It is important to note that the diagonal terms representing variances in the new estimates are smaller than those for either Example 1 or Example 2.

REFERENCES

1. Harkness, William, "The Solar Parallax and its Related Constants, Including the Figure and Density of the Earth," Washington Observations for 1885, Appendix III, 1891.

2. DeSitter, W., and D. Brouwer, "On the System of Astronomical Constants," Bull. Astron. Inst., Netherlands, Vol. 8, No. 307, July 8, 1938, pp. 213-231.

3. Clemence, G.M., "On the System of Astronomical Constants," Astron. J., Vol. 53, 1948, p. 169.

4. Herrick, S., R.M.L. Baker, Jr., and C.G. Hilton, "Gravitational and Related Constants for Accurate Space Navigation," UCLA Astron. Papers, Vol. 1, No. 14, 1958, p. 297.

5. De Vaucouleurs, G., The Astronomical Unit of Distance: Solar Parallax and Related Constants, The RAND Corporation, RM-2944-NASA, December 1961.

6. Arley, N. and K.R. Buch, Probability and Statistics, John Wiley and Sons, Inc., London, 1950.

7. Brown, Duane, A Matrix Treatment of the General Problem of Least Squares Considering Correlated Observations, Ballistic Research Laboratories, Report No. 937, May 1955.

8. Linnik, Yu.V., Method of Least Squares and Principles of the Theory of Observations, Pergaman Press, 1961.

9. Kaula, W.M. and I. Fischer, Army Map Service Technical Report No. 27 (Part I. Methods), U. S. Army World Geodetic System 1959, November 1959.

10. Cramer, Harald, Mathematical Methods of Statistics, Princeton University Press, 1951.

11. Deming, W.E., Statistical Adjustment of Data, John Wiley and Sons, Inc., London, 1948.

12. Madansky, Albert, and Allan Marcus, Statistical Analysis of Astronomical Unit Estimates, The RAND Corporation, RM-3940

13. Kocher, George, and Dean Jamison, A Deep-Space Triangulation Probe to Determine the Astronomical Unit, The RAND Corporation RM-4014-NASA, January 1964.